

3D Face Analysis for Facial Expression Recognition

Ahmed Maalej

LIFL, Université de Lille 1, France.
ahmed.maalej@telecom-lille1.eu

Anuj Srivastava

Departement of Statistics, Florida
State University, USA.

Boulbaba Ben Amor, Mohamed Daoudi

LIFL, Université de Lille 1, France.
Institut TELECOM/TELECOM Lille 1, France.

Stefano Berretti

Dipartimento di Sistemi e Informatica,
University of Firenze, Italy.

Abstract

In this paper, we investigate the person-independent 3D facial expression recognition. A 3D shape analysis is applied on local regions of 3D face scan. The correspondent regions of different faces under different expressions, are extracted and represented by a set of closed that capture their shapes. A framework is applied to quantify the deformations between curves and compute the geodesic length (or distance) that separates them. These measures are employed as inputs to a commonly used classification techniques such as AdaBoost and Support Vector Machines (SVM). A quantitative evaluation of our novel approach is conducted on the publicly available BU-3DFE database.

1 Introduction and recent advances

Facial expression analysis and recognition is an active and challenging research topic in computer vision, which guided researchers to develop applications in the field of *Human-Machine Interaction*. To study the way in which humans synthesize their emotions, we need to start to study the anatomical and biomechanical foundation of facial expression morphology. The pioneering work of *Ekman et al.* [1] introduced in the early 70s the *Facial Action Coding System*. FACS is anatomically based and allows the reliable coding of any facial action in terms of the smallest visible units of muscular activity. These units are called Action Units (AU), each referred to by a numerical code. AU combinations allow humans to build up prototypical facial expressions of basic emotions which can be categorized to six classes : *happiness, fear, sadness, surprise, disgust, and anger*. In the last few years, 3D face shape analysis has emerged as a major research trend in face

recognition for its robustness to changes in lighting conditions and pose variations. However, one still needs to deal with intra-class variations related to facial expressions. In fact, serious deformations caused by facial expressions, sometimes difficult to model (open mouth problem), present a very challenging problem. Thanks to these deformations, the 3D facial shape presents an interesting way to study facial expressions. This new method has attracted researchers to conduct work in 3D expressions recognition. *Wang et al.* [2] proposed to extract and label primitive surface features, and derive their statistical distributions for selected regions on the face scan. A person-independent facial expression recognition evaluation is conducted on the BU-3DFE database with an overall recognition rate of 83.6%. In [4], *Tang et al.* proposed an automatic feature selection method, the features are normalized Euclidean distances between a set of 83 ground truth points. Using a regularized multi-class AdaBoost classification algorithm, they achieve recognition rate of 95.1% .

In this paper, we further investigate the problem of 3D facial expression recognition. We propose a novel approach based on shape analysis of facial surfaces. More specifically, we are interested in studying local regions that we will call *patches* on the face and quantify their deformations under expressions. Finally, to categorize a given expression, we use two of the commonly used classifiers AdaBoost and SVM. The remainder of this paper is organized as follows: In section 2, we describe how we define and extract the patches around selected landmarks from facial surfaces. In section 3, we present our shape analysis framework for the 3D curves comparison and its extension to 3D patches analysis. In section 4, we report the experimental results and quantitative evaluation using the AdaBoost and the SVM algorithms for unbalanced classification problems. Finally, conclusion and future work are presented in section 5.

2 3D Facial Surfaces Representation

To recognize people, *Samir et al.* [5] represented facial surfaces by an indexed collection of 3D closed curves. These facial curves were characterized by one reference point, that is the nose tip, and were extracted over all the face surface. Then, the authors use a shape analysis framework to compare faces. In contrast, the proposed approach is oriented to study locally facial patches. Thus, instead of starting from only one reference point, we choose multiple reference points $\{r_l\}_{1 \leq l \leq N}$, and we extract curves $\{c_\lambda^l\}_{1 \leq \lambda \leq \lambda_0}$ over the face regions centered in these points. In the Binghamton University 3D facial expression (BU-3DFE) database, collected by *Yin et al.* [3], there is an MPEG4 specification of 83 facial feature points marked on all 3D facial expression cropped models of the 3D static database. The main purpose of these points is to provide spatial references to key positions on a human face, and they are chosen to reflect the best facial anatomy and biomechanical behavior of face's regions. We consider these feature points as reference points to locate patches and extract curves. Let r_l be a reference point corresponding to a specific patch P_l on the facial surface, denoted by S . Then, we define an Euclidean distance function $d : S \times S \rightarrow \mathbb{R}_+$, from r_l to any point p of the facial surface S , with r_l and $p \in \mathbb{R}^3$, so that $d(r_l, p) = \|p - r_l\|$. Using this function, one can define the facial curves as the level sets of $d(r_l, \cdot)$:

$$c_\lambda^l = \{p \in S | d(r_l, p) = \lambda\} \subset S, \lambda = \lambda_0. \quad (1)$$

Each c_λ^l is a closed curve, consisting of a collection of points belonging to S and situated at an equal distance λ from r_l . The Figure 1 illustrates examples of extracted level curves over different regions of a face scan after surface smoothing.

3 Framework for 3D Shape Analysis

Having these patches around some landmarks on the facial surfaces, study of their shapes can learn us about the localization of muscle deformations on the face and consequently its emotional state. Under the general hypothesis that the subjects on the database smile, or convey any other expression, the same way, or more properly the local regions that take place in the specific expression may undergo practically the same process of dynamical deformations. Consequently we expect that the value of the distance, computed by the framework, between two patches associated to the same expression is smaller than the one computed between two patches belonging to two different expressions. In the following

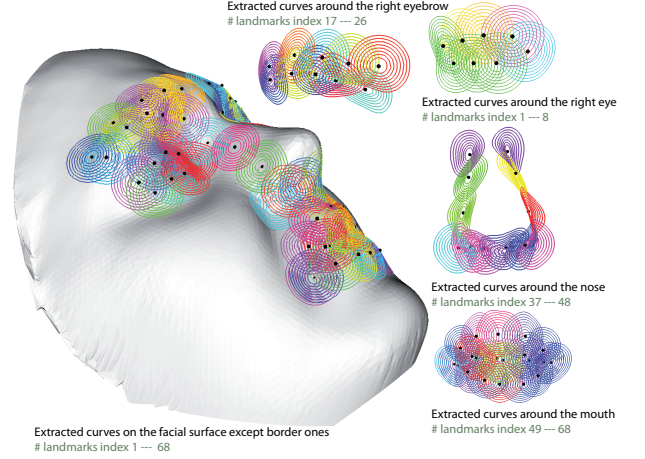


Figure 1. Curves extraction for facial regions representations centered on multiple reference points.

sections, we first present our framework for 3D curves analysis, then, we extend the ideas to capture shapes of 3D Patches.

3.1 3D Curve Shape Analysis

Thanks to the developed framework by *Joshi et al.* [6], we are able to perform curve-shape analysis to measure the distance between curves and to compute the geodesic path, which represent the optimal deformation from one shape to another. To find geodesic paths between patches, one needs to find the geodesic paths between correspondent curves. The framework presents a square-root velocity function (SRVF) denoted by q , that captures the shape of a given curve β . For more details about the proposed curve shape analysis framework, reader can refer to [6]. In the next we will use the notation $d(\beta_1, \beta_2)$ to denote the geodesic distance, or the length of the geodesic in S , between the two curves β_1 and β_2 .

3.2 3D Patches Shape Analysis

Now we extend ideas developed in the previous section for analysing shapes of curves to the shapes of patches. As mentioned earlier, we are going to represent a patch of a facial surface S with an indexed collection of the level curves of the d function (euclidean distance from the reference point r_l). That is, $P_l \leftrightarrow \{c_\lambda^l, \lambda \in [0, \lambda_0]\}$, where c_λ^l is the level set associated with $d = \lambda$. Through this relation, each patch has

been represented as an element of the set $\mathcal{C}^{[0, \lambda_0]}$. In our framework, the shapes of any two patches are compared by comparing their corresponding level curves. Given any two patches P_l^1 and P_l^2 , and their level curves $\{c_\lambda^{l_1}, \lambda \in [0, \lambda_0]\}$ and $\{c_\lambda^{l_2}, \lambda \in [0, \lambda_0]\}$, respectively, our idea is to compare the patches curves $c_\lambda^{l_1}$ and $c_\lambda^{l_2}$, and to accumulate these differences over all λ . More formally, we define a distance d_s given by:

$$d_s(P_l^1, P_l^2) = \int_0^L d(c_\lambda^{l_1}, c_\lambda^{l_2}) d\lambda. \quad (2)$$

It has been shown in [5] that this distance is actually a proper distance between the elements of the $\mathcal{C}^{[0, \lambda_0]}$. In addition to the distance $d_s(P_l^1, P_l^2)$, which is useful in biometry and other classification experiments, we also have a geodesic path in $\mathcal{C}^{[0, \lambda_0]}$ between the two points represented by S^1 and S^2 . This geodesic corresponds to the optimal elastic deformations of facial curves and, thus, facial surfaces from one to other. Shown in Figure 2 are examples of such geodesic paths between given facial surfaces. For both sides of each geodesic path, the first surface is reconstructed from the set of 3D curves extracted from the patch indicated on the nearby face scan, between the two surfaces we illustrate the computed optimal elastic deformations. The top two rows correspond to geodesic paths that involve the same source, a person under a happy expression (right), and two different targets (left), the same person under a surprise state (first row) and a different person with happy expression (second row). We note that the geodesic distance between the same face under different expressions is greater than the one that separates the two different persons conveying the same expression.

4 Experimentations and evaluations

In this section, we present preliminary results of our approach, and an evaluation of its performance on the Binghamton University 3D facial expression (BU-3DFE) database, created by Yin et. al. [3] for facial behavior research. In the BU-3DFE there is 100 subjects, partitioned into two genders 56 females and 44 males, with variety of ethnicity and racial ancestries. Each subject was asked, while capturing 3D data, to perform a neutral facial expression as well as the six prototypic facial expressions as categorized by most studies *Ekman* [1]. Each facial expression has 4 level of intensities low (01), middle (02), high (03) and highest (04), except for neutral faces which have only one intensity level (00). We conduct our experiments on 20 subjects, each one presents the 6 facial expressions (*ha*, *an*, *di*, *fe*, *su* and *sa*) of the highest level (04). Then we extract

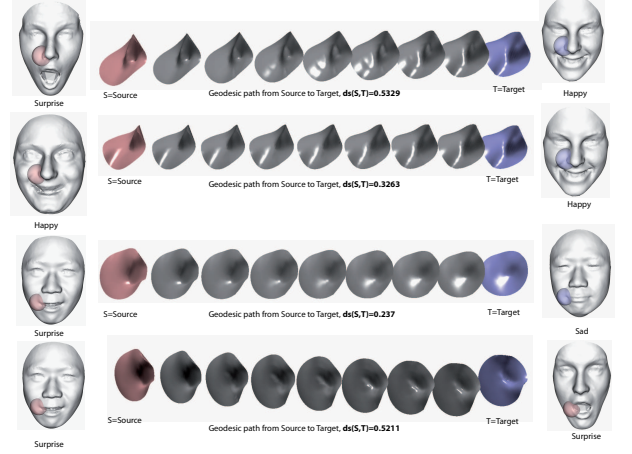


Figure 2. Illustration of some geodesic paths between patches.

patches centered on a finite number of feature points for each face. Based on the the hypothesis of symmetric property of the human face, we optimize our study by using only the feature points laying on half a face model (left part). The total number of the chosen feature points is 24 as illustrated, with black points, in Figure 3. After patches extraction, we compute a similarity

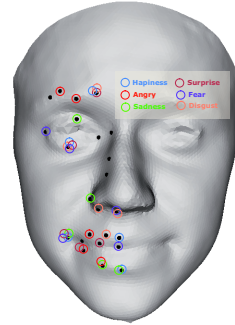


Figure 3. 24 Selected feature points for patch extraction.

matrix P with dimension 120×120 , the data in P are arranged to form six submatrices of dimension 20 each. For the first submatrix we have in rows the 20 corresponding patches of the 20 chosen persons relative to a happy expression, in columns we have the same persons under all the prototypic expressions, we can further decompose the submatrix into six blocks; the first 20 columns represent the patches of the same persons, in the same order as in rows, and with the same ex-

pression, while the second 20 columns represent the angry expression, the third 20 ones represent disgust, and so on till the sad expression. We end up with a matrix M of dimension $120 \times 120 \times 24$, where $M(i, j, k)$ is the distance that separates between two patches belonging to subjects i and j according to the same feature k . Thus, this matrix contains ha vs. ha comparisons, ha vs. \overline{ha} comparisons, an vs. an comparisons, an vs. \overline{an} comparisons, and so on. This matrix allows us to learn six binary classifiers for each classification ha/\overline{ha} , an/\overline{an} , di/\overline{di} , fe/\overline{fe} , su/\overline{su} , sa/\overline{sa} . We employ two commonly used classification techniques, AdaBoost and SVM. Both algorithms take as input a set of training samples $(x_1, y_1), \dots, (x_N, y_N)$ where each x_i ($i = 1, \dots, N$) is an sample that belongs to some domain or instance space X , and each label y_i is a boolean value that belongs to the domain $Y = \{-1, +1\}$, indicating whether x_n is positive or negative sample. In our case we consider the submatrix P_{ha} of dimension 20 to classify the happy vs. non-happy expression, we label positively the block 20 and the rest is labelled negatively. Half of this submatrix is retained for training (10) the classifiers while the other half is used for testing. The results of binary classification using AdaBoost, linear-SVM, Polynomial-SVM and Radial Basis Functions RBF-SVM, are presented in Figure 4.

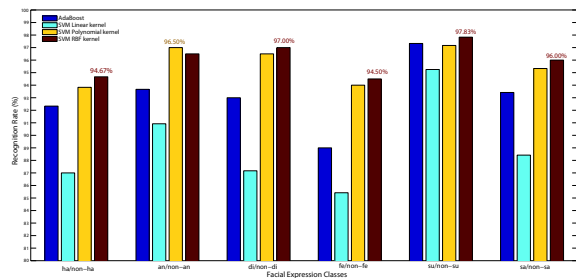


Figure 4. Results of AdaBoost and SVM performances for facial expression classification.

Although RBF-SVM gives better classification performances, the application of the AdaBoost algorithm enrich our face analysis in the way that it permits to identify most relevant patches for the expressions classification. In Figure 3 we labelled the six expressions with different colors and we marked with the appropriate color, the feature points that are highly involved, according to the affected weight across the AdaBoost test set, in characterizing the deformation under a given expression. The resulting average classification rates, using the mentioned techniques, are summarized in Table

| AdaBoost | SVM-Linear | SVM-Polynomial | SVM-RBF |
|----------|------------|----------------|---------------|
| 93.12% | 89.03% | 95.63% | 96.08% |

Table 1. Average classification rates.

1, the highest one is achieved with RBF-SVM rate equal to 96.08%.

5 Conclusion

In this paper we investigate the problem of 3D face analysis for person independent facial expression recognition. We propose a novel approach based on the measure of inter-face expressional variations over multiple selected regions on the 3D face model. These curves-based representation is applied to capture the deformation between correspondent patches on different faces under different expressions. A quantification of this deformation in terms of distance (or length) of the geodesic paths is computed based on the Riemannian framework, these measures are then used as samples for AdaBoost and SVM binary classifiers. Since the proposed approach evaluate inter-class variations, it can be extended to perform identity recognition. As future work, we intend to enlarge our dataset to perform multiclass facial expression recognition.

References

- [1] P. Ekman, W. V. Friesen, and J. C. Hager. *The Facial Action Coding System (FACS): Manual*. Palo Alto, CA: Consulting Psychologists Press, 1978.
- [2] J. Wang, L. Yin, X. Wei and Yi. Sun. *3D Facial Expression Recognition Based on Primitive Surface Feature Distribution*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1399-1406, 2006.
- [3] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. *A 3D Facial Expression Database For Facial Behavior Research*. 7th IEEE International Conference on Automatic Face and Gesture Recognition (FG2006), pp.211-216.
- [4] H. Tang and T.S. Huang. *3D facial expression recognition based on automatically selected features*. Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1-8, June 2008.
- [5] C. Samir, A. Srivastava, M. Daoudi and E. Klassen. *An Intrinsic Framework for Analysis of Facial Surfaces*. International Journal of Computer Vision, vol. 82, no. 1, pp. 80-95, 2009.
- [6] S. Joshi, E. Klassen, A. Srivastava and I. H. Jermyn. *A Novel Representation for Efficient Computation of Geodesics Between n-Dimensional Curves*. IEEE Computer Vision and Pattern Recognition (CVPR), 2007.